

Sur-Rebuttal Report of Michael A. Campion, Ph.D.,

in

Chen-Oster v. Goldman Sachs

(10 Civil 6950 (AT) (JCF))

pending

in the U.S. District Court

for the Southern District of New York

Contains Confidential Information Pursuant To Court Order

Michael A. Campion, Ph.D.
Krannert Graduate School of Management
Purdue University
403 West State Street
West Lafayette, IN 47907-2056
Phone: 765-494-5909, Fax: 765-494-9658
Email: campionm@purdue.edu
Web Page: campion-services.com

May 7, 2014

Table of Contents

I.	Purpose of the Report.....	1
II.	Summary of Opinions	1
III.	Applicability of Uniform Guidelines and SIOP Principles to Performance Evaluation and Compensation	4
IV.	Uniform Guidelines	4
V.	SIOP Principles.....	6
VI.	Literature Review.....	7
VII.	Dr. Cascio Erred In Concluding That Goldman Sachs’ Performance Evaluation and Compensation Processes Are Inconsistent with the Job-Relatedness Concepts Found In the Uniform Guidelines and the SIOP Principles	11
VIII.	The Merits of A “Best Practices” Approach to Analyzing Performance Evaluation and Compensation	15
IX.	Goldman Sachs’ Use of 360 Feedback Scores Was Not Arbitrary.....	17

I. Purpose of the Report

1. Goldman, Sachs & Co. (“Goldman Sachs” or the “firm”) engaged the author through its counsel, Paul Hastings LLP, to review and respond to new matters contained in the rebuttal expert report of Wayne F. Cascio, Ph.D. (of January 28, 2014) (the “Rebuttal Report”) in the case of *Chen-Oster, et al. v. Goldman, Sachs, & Co., et al.* (10 Civil 6950 (AT)(JCF)).

2. This Sur-Rebuttal Report is limited to the opinions and topics raised in Dr. Cascio’s Rebuttal Report that were not included in his original Report (of October 30, 2013).

3. All analyses and conclusions in my previous report (the “Initial Report”)¹ and deposition (of December 30, 2013) are incorporated. All materials relied upon are included in the list of References attached to this Report.

II. Summary of Opinions

4. Dr. Cascio says in his Rebuttal Report — for the first time — that Goldman Sachs’ performance evaluation and compensation processes *must* be validated using the methodology set forth in the Uniform Guidelines on Employee Selection Procedures (the “Uniform Guidelines” or the “Guidelines”)² and the Principles for the Validation and Use of Personnel Selection Procedures (the “SIOP Principles” or the “Principles”).³ However, he does not refer to any provisions in those publications that discuss a methodology for validation of evaluation or compensation processes, or that could reasonably be applied directly to these

¹ Report of Michael A. Campion, Ph.D., on Performance Evaluation (360 Feedback and Quartiling) and Incentive Bonus Compensation in *Chen-Oster v. Goldman Sachs* (10 Civil 6950 LBS) pending in the U.S. District Court for the Southern District of New York (December 11, 2013).

² The Uniform Guidelines were published by the EEOC, Civil Service Commission, Department of Labor, and Department of Justice in 1978, and may be accessed at <http://www.uniformguidelines.com/uniformguidelines.html> (last visited May 7, 2014).

³ The Principles were published by the Society for Industrial and Organizational Psychology (“SIOP”) in 2003, and may be accessed at http://www.siop.org/_principles/principles.pdf (last visited May 7, 2014).

processes — the publications are silent on these subjects. He also suggests — again, for the first time — that the Goldman Sachs processes are not valid under the Guidelines and Principles, but he provides no analysis to support that conclusion. Dr. Cascio appears to be attempting to create an artificial criterion for validity, which he then claims in a wholly conclusory fashion that Goldman Sachs does not meet. There is no basis for this.

5. While the Guidelines and Principles can provide an important framework for considering the validity of an employer's performance evaluation and compensation processes, there is nothing in those documents to suggest that the methodology they describe (which relates to employer "selection" processes) can or should be applied directly to performance evaluation and compensation processes. The research literature supports that conclusion; there is very little support for the proposition that either performance evaluation or compensation processes must be validated using the methodology in the Guidelines and Principles (and, indeed, recognition that these resources are not applicable in these contexts).

6. Unlike Dr. Cascio, who in his Rebuttal Report has largely ignored the actual content of the Guidelines and Principles, I explained in my Initial Report how the concepts found in those publications can be useful in assessing the job-relatedness of the Goldman Sachs evaluation and compensation processes. I then demonstrated that the evaluation processes in particular are "content-valid" in part by using those validation concepts. *See, e.g.*, Tables 3-4 to Initial Report. In his Rebuttal Report, Dr. Cascio rejects my opinion in a conclusory way, but his Rebuttal Report offers no specific examples of how my approach deviates from the concepts articulated in Guidelines and Principles and, indeed, his Rebuttal Report largely ignores the specifics of my analysis.

7. Dr. Cascio's Rebuttal Report also suggests for the first time that a process cannot be valid under the Uniform Guidelines or the Principles if it produces a gender difference, but that contention turns the analysis upside down; validation of a selection procedure is not required *unless* that procedure produces an adverse impact, and the Uniform Guidelines expressly state that. The existence of a gender (or other protected characteristic) difference is what *triggers* application of the Guidelines and Principles (as and when they are otherwise applicable). But gender differences do not mean that the Guidelines and Principles have not been satisfied. Here, even assuming that gender differences exist (a subject addressed in the Expert Report of Michael P. Ward, Ph.D.), I reaffirm my prior analyses and evidence that the Goldman Sachs processes are job-related and therefore valid.

8. With respect to the concept of "best practices," Dr. Cascio's Initial Report acknowledged the central role of "best practices" in our field. *See* Initial Report at ¶ 75; § V, p. 6. In his Rebuttal Report, however, he reverses himself and says that best practices should play little or no role in assessing employment processes, and he criticizes me for taking them into account. His change of opinion is not justified. Especially where, as here, the methodology in the Guidelines and Principles cannot be directly applied, it is entirely appropriate to perform an additional analysis comparing the processes at issue to "best practices" based on an exhaustive review of the research and professional literature, as I did in my Initial Report. Indeed, Dr. Cascio has used this approach in his own articles.⁴

9. In his Rebuttal Report, Dr. Cascio opines for the first time that Goldman Sachs acted "arbitrarily" by using 360 Feedback ratings as one input into employment decision-making

⁴ Dr. Cascio has used counts of articles on topics as an indicator to evaluate importance in his own research. Cascio, W. F., & Aguinis, H. (2008). Research in industrial and organizational psychology from 1963 to 2007: changes, choices, and trends. *Journal of Applied Psychology*, 93(5), 1062-1081.

because, according to him, those ratings were “severely” compressed. *See* Rebuttal Report at ¶ 45. Dr. Cascio did not say what standard he used for concluding that Goldman Sachs’ 360 Feedback ratings were “severely” compressed or how he might have derived such a standard, and he apparently performed no independent work to determine whether the 360 Feedback scores are, in fact, unusually compressed. A review of the data available to him (*i.e.*, the actual distribution of scores at Goldman Sachs and results commonly observed in the research literature) would have shown that the dispersion of ratings given by raters in the Goldman Sachs 360 Feedback process is typical, adequate, and useful.

III. Applicability of Uniform Guidelines and SIOP Principles to Performance Evaluation and Compensation

10. Neither the Uniform Guidelines nor the Principles contain any reference to performance evaluation or compensation in their titles, tables of contents, or text, aside from the limited exceptions noted below. Given this absence, there is no basis for concluding—as Dr. Cascio now asserts—that the methodology set forth in the Guidelines and Principles can or should be directly applied to evaluation or compensation processes. In fact, they are rarely used for “validating” these two Human Resources processes. Nevertheless, as I showed at length in my Initial Report, the Guidelines and Principles do describe concepts that are important in assessing the validity of those processes. Those concepts may be helpful to some extent in assessing the job-relatedness of evaluation and compensation processes, but Dr. Cascio’s contention that these standards directly impose a benchmark that must be met is wrong.

IV. Uniform Guidelines

11. As the full title of the Guidelines — the “Uniform Guidelines on Employee *Selection* Procedures” — reflects, the publication’s focus is on *selection* decisions such as hiring and promotion, and not on performance evaluation or compensation. This focus is also evident

in the table of contents for the Guidelines, which exclusively refers to selection topics.⁵ The table of contents does not even mention compensation, and only references performance evaluation as one possible criterion to be used in a validation study for selection processes.⁶

12. The text of the Guidelines does not refer to compensation or the use of performance evaluations in compensation processes, and these processes are not included within the Guidelines' definition of applicable "employment decisions" (Section 2B):

These guidelines apply to tests and other selection procedures which are used as a basis for any employment decision. Employment decisions include but are not limited to hiring, promotion, demotion, membership (for example, in a labor organization), referral, retention, and licensing and certification, to the extent that licensing and certification may be covered by Federal equal employment opportunity law. Other selection decisions, such as selection for training or transfer, may also be considered employment decisions if they lead to any of the decisions listed above.⁷

⁵ See <http://www.uniformguidelines.com/uniformguidelines.html> (last visited May 7, 2014).

⁶ Performance evaluations are often used as criteria in validation studies to show the job-relatedness of selection procedures. For example, an employer might correlate a pre-hire test used for job applicants with subsequent job performance evaluation ratings as a means of demonstrating that the pre-hire test did, in fact, have criterion-related validity. That is, if the individuals passing the pre-hire test performed adequately in the job once hired, the performance evaluations could provide criterion-related validity evidence that the pre-hire test was job-related.

⁷ It was not until the EEOC published its "Questions and Answers To Clarify and Provide a Common Interpretation of the Uniform Guidelines on Employee Selection Procedures" a year later in 1979 that any reference was made to performance evaluations. Even then, however, no guidance was provided as to how the methodology in the Guidelines could be applied:

Q. Do the Guidelines apply only to written tests?

A. No. ***They apply to all selection procedures used to make employment decisions***, including interviews, review of experience or education from application forms, work samples, physical requirements, and ***evaluations of performance***.

See Adoption of Questions and Answers to Clarify and Provide a Common Interpretation of the Uniform Guidelines on Employee Selection Procedures. 44 Fed. Reg. 11996-12009 (1979). (emphasis added).

13. Although types of “selection” processes (such as hiring, promotion, and firing) are cited hundreds of times in the Uniform Guidelines and in the Questions and Answers that the EEOC published to clarify those Guidelines, job performance is referenced rarely, and even then only in a list of tools used to make selection decisions such as interviews, job applications, and physical requirements — things that are not employee selections in and of themselves.⁸ Nothing in the Uniform Guidelines indicates that performance evaluations or compensation processes need to be independently validated, and they provide no procedures for doing so.

V. SIOP Principles

14. Like the Guidelines, the full title of the Principles (the “Principles for the Validation and Use of Personnel *Selection* Procedures”) shows their focus on selection procedures, not performance evaluation and compensation processes. The table of contents for the Principles also reflects this emphasis; like that of the Guidelines, this table of contents has no reference to validating performance evaluation or compensation processes.⁹

15. The text of the Principles similarly indicates that they were not intended to guide validation of performance evaluations. For example, the definition of “personnel decisions” within the scope of the Principles does not include performance evaluations (p. 3):

Personnel decisions are employment-related decisions to hire, train, place, certify, compensate, promote, terminate, transfer, and/or take other actions that affect employment status.

⁸ See, e.g., Adoption of Questions and Answers to Clarify and Provide a Common Interpretation of the Uniform Guidelines on Employee Selection Procedures. (1979). 44 Fed. Reg. 11996-12009 (1979). ((1) referencing performance evaluations as “selection procedures used to make employment decisions (also cited by Dr. Cascio in Paragraph 15 of his Rebuttal Report); and (2) using performance evaluations as one predictor for selecting individuals in training programs or in probationary periods for permanent employment).

⁹ See http://www.siop.org/_principles/principles.pdf (last visited May 7, 2014). While this table of contents mentions “performance-related criteria” once, this is in the context of validating personnel selection practices.

While the Principles cite performance evaluations as a type of “selection procedure,”¹⁰ they do not address validation of evaluation processes or provide any methodology that might be used for that purpose.

16. As with the Guidelines, the Principles’ selection-related focus is confirmed by the selection-related terms throughout the document. The Principles contain only isolated references to terms relating to job performance. When performance evaluation is mentioned, it is identified as a validation criterion for *other* processes, not as a process that itself requires validation.¹¹ Indeed, the Principles caution administrators against using selection procedure information for purposes other than intended, and note: “[A]lthough selection procedure data may have some validity in determining later retention decisions, more potentially relevant measures such as performance ratings may be available.” (p. 57).

VI. Literature Review

17. A review of the research literature shows that the Uniform Guidelines and Principles are virtually never cited in the context of performance evaluation or compensation processes; they are almost always cited in the context of selection.

¹⁰ “Selection procedures refer to any procedure used singly or in combination to make a personnel decision including, but not limited to, paper-and-pencil tests, computer-administered tests, performance tests, work samples, inventories (e.g., personality, interest), projective techniques, polygraph examinations, individual assessments, assessment center evaluations, biographical data forms or scored application blanks, interviews, educational requirements, experience requirements, reference checks, background investigations, physical requirements (e.g., height or weight), physical ability tests, appraisals of job performance, computer-based test interpretations, and estimates of advancement potential.” (p. 3)

¹¹ See, e.g., the “scope” passage quoted by Dr. Cascio in Paragraph 13 of his Rebuttal Report (p. 3); the observation that there is no “true score” against which performance evaluation results can be compared (*i.e.*, validated) (p. 33) — an important point described at page 10 of this Report; and the section describing the utility of selection scores as one tool in implementing a valid reduction-in-force process (pp. 56-57).

18. In response to the contentions Dr. Cascio makes in his Rebuttal Report regarding the applicability of the Guidelines and Principles to performance evaluation and compensation processes, I performed a literature search using PsycInfo Citation Search, which identifies all research publications citing a given document. The database contains peer-reviewed research in Industrial and Organizational Psychology, Human Resources, Organizational Behavior, and other fields related to management. (It does not include non-peer-reviewed publications in practitioner and public press articles, newspapers, electronic media, and related sources.)

19. The literature search for the Principles yielded 37 articles and book chapters (all are included in the list of References attached to this Report). All refer to selection; *none* refers to performance evaluation or compensation.

20. The literature search for the Uniform Guidelines yielded 148 articles and book chapters (all are included in the list of References attached to this Report). *None* of the citing publications referred to compensation. Only 4 publications (3%) referred to performance evaluation, and they did so in distinguishable contexts.¹²

¹² Norton, Balloun, and Konstantinovich (1980) states that the Guidelines “do not directly address the validity of supervisory ratings when used for promotion decisions” (p. 380) and therefore that these ratings are “inapplicable to [the promotion] situation” (p. 377). *See* Norton, Balloun, and Konstantinovich (1980). The soundness of supervisory ratings as predictors of managerial success. *Personnel Psychology*, 33(2), 377-388. Similarly, Wingate, Thornton, McIntyre, and Frame (2003) state that the Guidelines do not apply to reductions in force (p. 88), even though their analyses show that reductions often consider job evaluation information. *See* Wingate, Thornton, McIntyre, and Frame (2003). Organizational downsizing and age discrimination litigation: The influence of personnel practices and statistical evidence on litigation outcomes. *Law And Human Behavior*, 27(1), 87-108. Aguinis and Smith (2010) referred to potential biases in performance evaluations, but they did not assert that the Guidelines should be used to validate performance evaluation processes. *See* Aguinis and Smith (2010). Balancing adverse impact, selection errors, and employee performance in the presence of test bias. In J. L. Outtz (Ed.), *Adverse impact: Implications for organizational staffing and high stakes selection* (pp. 403-423). New York, NY US: Routledge/Taylor & Francis Group. Finally, Bernardin, Beatty, and Jensen (1980) referred to the Guidelines in the context of discussing the use of student evaluations in making university employment decisions, but provided no

21. In a few instances, researchers have considered the applicability of the Uniform Guidelines to performance evaluation, and the consensus is that they do not apply. An article written by Barrett and Kernan (1987)—the first article to state the viewpoint that performance evaluation need not be validated under the Uniform Guidelines—concluded:

The courts have not referred to the Uniform Guidelines as a standard which should be followed in evaluating a performance appraisal system in Title VII cases, nor do they require that a performance appraisal be considered a content valid test or that there be an independent empirical verification of the performance appraisal process.¹³

22. To evaluate professional consensus in the years following this article, I conducted a search to identify all research articles and books citing Barrett and Kernan under the assumption that those discussing the issue would cite this important article. The review yielded 34 publications (all are included in the list of References attached to this Report): (a) the majority (21 publications, or 62%) do not express an opinion on the application of the Uniform Guidelines to performance appraisals; (b) 10 (29%) appear to agree that the Guidelines do *not* apply in this context; (c) two (6%) appear to believe the Guidelines do apply; and (d) one (3%) simply notes the issue. Dr. Cascio did not attempt to marshal evidence from the literature in our field to support his new assertion that the Uniform Guidelines must be used to validate performance evaluations, and, in fact, the publications in our field show a lack of support for that assertion.¹⁴

particulars as to how the Guidelines might be used for that purpose. *See* Bernardin, Beatty, and Jensen (1980). The new Uniform Guidelines on Employee Selection Procedures in the context of university personnel decisions. *Personnel Psychology*, 33(2), 301-316.

¹³ Barrett, G. V., & Kernan, M.C. (1987). Performance appraisal and terminations: A review of court decisions since *Brito v. Zia* with implications for personnel practices. *Personnel Psychology*, 40(3), 489-503.

¹⁴ Like Barrett and Kernan, the articles reviewing court cases published *prior* to Barrett and Kernan also found no support for requiring validation. *See, e.g.,* Faley, R.H., Kleiman, L.S., &

23. Dr. Cascio himself authored one article citing Barrett and Kernan, and his article strongly implies agreement with the view that the Uniform Guidelines do not apply to performance evaluation. Dr. Cascio noted the frequent contention

by plaintiffs [in discrimination litigation] that performance appraisals also must be validated. This contention also has not been upheld by the courts (Barrett & Kernan, 1987). In fact there is probably a trend for the courts to understand that not all tests or qualifications for a job have to be “valid” (*Aguilera v. Cook County*, 1985).¹⁵

24. One key reason that the Uniform Guidelines and Principles do not require validation of performance evaluation processes is alluded to above: the Principles state that there is no “true score” for performance evaluations (p. 33). This means there can be no measure of job performance against which to validate or confirm the ratings received in a performance evaluation. To ensure that a performance evaluation is actually measuring “true” performance, one would have to have some other measure of what the “true” job performance actually was so that the performance evaluation results could be compared. Because the performance evaluation is, itself, the measure of employee performance, there is no “true score” against which the evaluation results can be validated.

Lengnick-Hall, M.L. (1984). Age discrimination and personnel psychology: A review and synthesis of the legal literature with implications of future research. *Personnel Psychology*, 37, 327-350; Field, H.S., & Holley, W.H. (1982). The relationship of performance appraisal system characteristics to verdicts in selected employment discrimination cases. *Academy of Management Journal*, 25, 392-406; and Kleiman, R.L., & Durham R.L. (1981). Performance appraisal, promotion and the courts: A critical review. *Personnel Psychology*, 34, 103-121.

¹⁵ Cascio, W. F., Alexander, R. A., & Barrett, G. V. (1988). Setting cutoff scores: Legal, psychometric, and professional issues and guidelines. *Personnel Psychology*, 41(1), 1-24.

VII. Dr. Cascio Erred In Concluding That Goldman Sachs' Performance Evaluation and Compensation Processes Are Inconsistent with the Job-Relatedness Concepts Found In the Uniform Guidelines and the SIOP Principles

25. The foregoing shows that Dr. Cascio was incorrect when he suggested for the first time in his Rebuttal Report that the Uniform Guidelines and the SIOP Principles must be directly applied in order to validate the performance evaluation and compensation processes at Goldman Sachs. But as I noted in my Initial Report, these materials can be used as tools in demonstrating key facets of “job-relatedness” that, in turn, can usefully be applied in assessing those processes for validity.

26. Although Dr. Cascio claimed (for the first time) in his Rebuttal Report that the processes at issue here are not valid under the Guidelines or Principles (*see generally* Rebuttal Report at §§ 2, 3), he never identified any provisions of those documents that might be said to undermine the validity of Goldman Sachs processes. Rather, while generally concluding that “Goldman Sachs’ performance evaluation and compensation-setting systems . . . are not valid” (Rebuttal Report at § 3), Dr. Cascio only mentions the Guidelines and Principles in support of his position that the firm should have sought, but allegedly did not seek, to “validate” these practices. *See, e.g.*, Rebuttal Report at ¶ 89. In short, Dr. Cascio offers an opinion on the validity of the Goldman Sachs processes based on the Uniform Guidelines and Principles, but he fails to “show his work.” His opinion on this point is merely an empty assertion, made without any substantiation.

27. Dr. Cascio’s Rebuttal Report is also almost entirely silent with respect to the specifics of the extensive work I described in my Initial Report drawing validation lessons from the Guidelines and Principles and applying them to the record in this case.

28. Similarly, Dr. Cascio's Rebuttal Report largely ignores the fact that both of the Goldman Sachs processes at issue here were subject to extensive review by job analysis studies conducted by reputable outside consulting firms in our field. At his deposition, Dr. Cascio acknowledged that he had no quarrel with the quality or content of that work. *See* Deposition of Wayne Cascio at 57:23-58:21; 59:10-24.

29. Thus, Dr. Cascio's Rebuttal Report does not meaningfully address, much less undermine, my conclusion in my Initial Report that these Goldman Sachs processes meet the requirements to demonstrate content validity as described in both the Uniform Guidelines and Principles. In brief:

- a. The 360 Feedback process was developed based on a job analysis, it measures a sample of job behavior, and the behavior measured is observable and important to the jobs at issue. The 360 Feedback process is inherently job-related because it asks raters to assess the job performance of the rated employees. Its validity is nearly tautological.
- b. The quartiling and compensation processes rely on many of the same performance measures used in the 360 Feedback process, but also incorporate important job performance outcomes through productivity metrics and other job-related factors such as potential for future growth and commitment to firm citizenship principles such as diversity.
- c. During the relevant period, Goldman Sachs retained outside research firms to perform multiple, extensive job analysis studies, and the results were used to develop the performance evaluation (360 Feedback) process and to confirm that the compensation factors were job-related, thus supporting content validity. *See, e.g.*, Initial Report ¶¶ 39-43; 70-76; Tables 5-8.
- d. To evaluate the extent of job-relatedness directly, I included in my Initial Report comparisons between the job analysis results and the factors in the 360 Feedback and compensation (quartiling) processes, thus further supporting content validity. There is a one-for-one link between the competencies identified by the job analysis and the criteria used in the 360 Feedback process surveys.

30. Finally, Dr. Cascio suggests in his Rebuttal Report that an employment decision-making process cannot be valid if it produces gender differences (or differences adverse to some other protected category). This assertion is curious, because the Uniform Guidelines on which

he now seeks to rely expressly provide that they are applicable *only* when “the use of [a] selection procedure . . . has an adverse impact on the hiring, promotion, or other employment or membership opportunities of members of any race, sex, or ethnic group[.]” *See* Uniform Guidelines § 3(A). Gender differences are not an indication of improper implementation or lack of validity; they are merely an indication that (where applicable) the Uniform Guidelines should be consulted to *assess* validity.

31. Dr. Cascio makes the following statement in his Rebuttal Report (§ 31):

“In my view, the outcomes of a process, in this case, the extent to which performance reviews and pay disproportionately disadvantage female employees, are . . . *relevant to judgments about validity . . .*” (emphasis added)

32. This statement is erroneous. Differences in outcomes are *not* relevant to judgments about validity. An employment process is valid if it is job-related, *regardless of outcomes*. As an extreme example, imagine a hiring process that was based exclusively on the month in which the applicant was born — giving candidates born early in the year preference to those born later. This preference for applicants born in January could not be considered “valid,” because the decision-making criterion has no bearing on work tasks — *i.e.*, it is not “job-related.” But it almost certainly would generate no gender differences, because neither gender can lay claim to a disproportionate percentage of January births. Conversely, a decision-making criterion that *does* result in gender differences (*e.g.*, requiring an advanced degree in mathematics or computer science) may well be both valid — *i.e.*, job-related — *and* the source of a gender difference in outcomes.

33. Finally, in his Rebuttal Report, Dr. Cascio offered an alternative method for validating the Goldman Sachs performance evaluation and compensation processes. He offered alternate means for Goldman Sachs to “subject its practices (*e.g.*, the 360-degree performance

review as implemented, the forced ranking, and the compensation-setting) to the rigor of formal reliability and validation studies.” In particular, Dr. Cascio opined that

[o]ne . . . strategy would be to [determine] *[i]f individual performance really does predict important outcomes such as compensation . . . decisions [If that were the case] then presumably there should be a high degree of relationship (correlation) between performance ratings (360-degree reviews or Quartile assignments) and bonuses Criterion-related validity is certainly feasible under these circumstances.*

Rebuttal Report at ¶ 25 (emphasis added).

34. I was not retained to perform any statistical analyses in this case, and do not offer any opinion on any statistical matter in this Report, but I note that the Initial Report of Goldman Sachs’s retained expert on statistics, Dr. Michael P. Ward, directly addresses the correlation between 360 Feedback scores and manager quartile placement. He concluded that there was “a dramatic statistically significant positive correlation between 360 score and manager quartile assignments within a Business Unit for employees with the same corporate title (*e.g.*, for Vice Presidents in FICC Americas Sales).” See Expert Report of Michael P. Ward, Ph.D. in the Matter of *Chen-Oster et al. vs. Goldman, Sachs & Co. and The Goldman Sachs Group, Inc.* at p. 15. I am unaware of any contrary statistical evidence submitted by any expert in this matter.

35. Similarly, I am informed that in his Sur-Rebuttal Report to be served contemporaneously with this Report, Dr. Ward analyzes the correlation between 360 Feedback scores and manager quartile placement, on the one hand, and compensation results on the other. Although the impact of those two performance measures varies by Divisions, job levels (*i.e.*, Associates and Vice Presidents) and lateral/non-lateral status, there is overall a “a highly statistically significant impact on the pay of all employees.”

36. As Dr. Cascio suggested, these correlations provide independent validity evidence that affirms the conclusions I reached in my Initial Report regarding the validity of the Goldman Sachs evaluation and compensation processes.

VIII. The Merits of A “Best Practices” Approach to Analyzing Performance Evaluation and Compensation

37. Although Dr. Cascio’s Initial Report acknowledges the important role served by “best practices” in our field (*see* Initial Report at ¶ 75; § V, p. 6), he reverses course and claims in his Rebuttal Report that it was inappropriate for me to use best practices as one tool among several for assessing Goldman Sachs’ evaluation and compensation processes. That criticism is unfounded.

38. Because the Uniform Guidelines and Principles are not directly applicable to validating performance evaluation and compensation processes, I performed an additional analysis, comparing the Goldman Sachs processes to “best practices” based on an exhaustive review of the research and professional literature. This approach has many merits.

39. First, it makes the analysis more current. The Uniform Guidelines were published in 1978 and thus do not reflect advances in the field over the last 35 years. Likewise, the Principles are more than 10 years old (published in 2003). A comprehensive review of the practices recommended in the literature on the topics relevant to the case provides a more current and exhaustive standard of comparison than either the Guidelines or the Principles. (As discussed above, neither the Guidelines nor the Principles deal with the topics relevant to the case—performance evaluation and compensation processes—in any meaningful way.)

40. As Dr. Cascio recognized in his Initial Report, the use of best practices as one standard against which to evaluate organizations is a common approach in the field of Human Resources. Organizational leaders want to know how they compare to the best. It is also a

common teaching tool in Human Resources, and thus is an assessment tool I use in my non-litigation work.¹⁶ Best practices are intended to provide a high standard; they are a useful point of comparison.

41. As part of his new criticism of the “best practices” approach, in his Rebuttal Report, Dr. Cascio disparages articles that reflect practical experience in the field rather than results obtained from a research study. Articles in professional publications emphasizing the practical experiences of professionals in the field, however, should be given some weight as indicators of effective practice, and that is why I included them in my Initial Report. Not every topic has been the subject of a research study for many reasons (*e.g.*, not of theoretical interest, not necessary to conduct a study to know what is best, etc.). Listening to the input of practitioners is important to understanding organizational phenomena, a fact that Dr. Cascio has acknowledged in his own publications:

[I]n modeling the effects of contextual factors that might contribute to the prediction of some organizational outcome, the input of practitioners or managers with first-hand experience and in-depth knowledge of an organization is, in our opinion, even more important if the research is to demonstrate ecological validity (accurately represent the pattern of relationships between employees and their organizational environments).¹⁷

42. Moreover, none of the practices cited in my Initial Report (shown in Tables 9 and 11 of that Report) was supported only by “practically oriented” professional literature; all of the practices were also supported by research literature.

¹⁶ In teaching, I use a textbook that explicitly focuses on “best practices” (Hollenbeck, J. R., Gerhart, B., & Wright, P. M. (2004). *Fundamentals of human resource management* (4th ed). New York: McGraw-Hill Irwin.).

¹⁷ Cascio, W. F., & Aguinis, H. (2008). Research in industrial and organizational psychology from 1963 to 2007: changes, choices, and trends. *Journal of Applied Psychology*, 93(5), 1077.

43. Finally, I was exhaustive, rather than selectively partisan, in the “best practices” materials I cited. Some of the articles I cited do not support the best practices I identified. Nonetheless, I included *all* of the literature I found, both supportive and not, to show a complete and unbiased approach. (This is what scientists are supposed to do when performing work using the scientific peer review process.)¹⁸ The great preponderance of the literature supports the “best practices” I identified in every case.

IX. Goldman Sachs’ Use of 360 Feedback Scores Was Not Arbitrary

44. In his Rebuttal Report Dr. Cascio also offers the opinion — for the first time — that it was “arbitrary” for Goldman Sachs to rely on scores obtained from the 360 Feedback process as a factor in making employment decisions because, he says, those scores were “severely” compressed. *See* Rebuttal Report at ¶ 45. It appears, however, that Dr. Cascio did not review the available data to determine for himself whether the scores were unusually compressed, and he cites no research literature for any standard or benchmark for determining when to classify scores as too compressed to be useful.

45. Rather than comparing the actual data to a standard derived from the research literature, it appears that Dr. Cascio merely relied on an email from a Goldman Sachs’ Human Resources executive, who explained that certain changes about to be made to the 360 Feedback process were intended to address perceived score compression. In my opinion, in the absence of an assessment of the data and a review of the pertinent literature, this purported “admission” from a Goldman Sachs employee is not a valid basis on which to offer and substantiate an expert opinion.

¹⁸ Campion, M. A. (1993). Article review checklist: A criterion checklist for reviewing research articles in applied psychology. *Personnel Psychology*, 46, 705-718.

46. People generally are either reluctant to give, or incapable of giving, adverse ratings in a performance evaluation setting — reluctant because they do not want to give critical feedback or incapable because they suffer from leniency bias. This frustrates the employers' attempt to get a wide variation in ratings.

47. This fundamental challenge with performance evaluations has been long recognized in the research, including in Dr. Cascio's own research as far back as 1977,¹⁹ and it is described in most textbooks (including Dr. Cascio's own textbook).²⁰ Nearly 25 years ago, a review of the performance appraisal literature concluded:²¹

It is common for 60 to 70% of an organization's workforce to be rated in the top two performance levels. This could reflect either actual outstanding performance on the part of the organization's workforce, or it could be indicative of leniency bias. Because the phenomenon is surprisingly constant across organizations, and it is unlikely that all organizations have predominately outstanding employees, the distributions probably reflect the latter. As anecdotal evidence and researchers' concern about leniency have suggested, it appears that the norm in U.S. industry is to rate employees at the top end of the scale. Skewed performance distributions not only exist, but are common.

48. The usual solution offered by those in my field to obtain more candid assessments is to use techniques such as managed distributions, which is what Goldman Sachs has done in its quartiling exercise. Dr. Cascio has criticized that approach in this case, but he recommended it in his textbook:

¹⁹ Cascio, W. F., & Valenzi, E. R. (1977). Behaviorally anchored rating scales: Effects of education and job experience of raters and ratees. *Journal of Applied Psychology*, 62(3), 278.

²⁰ Cascio, W. F., & Aguinis, H. (2005). *Applied psychology in human resource management* (6th ed.). Upper Saddle River, NJ: Pearson Prentice-Hall (see p. 97).

²¹ Bretz, R. D., Milkovich, G. T., & Read, W. (1992). The current state of performance appraisal research and practice: Concerns, directions, and implications. *Journal of Management*, 18(2), 321-352.

Leniency and severity biases can be controlled or eliminated in several ways [for example] by allocating ratings into a forced distribution, in which rates are apportioned according to an approximately normal distribution”²²

49. The email on which Dr. Cascio relies explained why Goldman Sachs was going to increase the range of available scores for the 2010 evaluation cycle by replacing its 5-point rating scale with a 9-point scale. Research in a variety of fields supports this sort of change, showing that use of a broader rating scale usually increases the variance of the scores.²³ Thus, this change was a reasonable effort to increase score differentiation.

50. But that does not mean the 5-point scale is too compressed to be useful. The most common scale for a 360 review reflected in the literature is the 5-point scale. Based on interviews with 211 companies in 30 industries, the 3D Group (2013) found that “[t]he 5-point scale has become the industry standard with over 85% of companies using this type scale.”²⁴

51. Rather than relying on one email to form my opinion on the “compression” issue raised by Dr. Cascio in his Rebuttal Report, I performed two inquiries: (a) I looked at the actual ratings data to determine the degree of dispersion or compression in the ratings; and (b) I researched the professional literature to determine whether “compression” existed at Goldman

²² Cascio, W. F., & Aguinis, H. (2005). *Applied psychology in human resource management (6th ed.)*. Upper Saddle River, NJ: Pearson Prentice-Hall (see p. 97).

²³ Examples include: Krosnick, J. A., & Fabrigar, L. R. (1997). Designing rating scales for effective measurement in surveys. *Survey measurement and process quality*, 141-164; Preston, C. C., & Colman, A. M. (2000). Optimal number of response categories in rating scales: reliability, validity, discriminating power, and respondent preferences. *Acta psychologica*, 104(1), 1-15; and Cook, D. A., & Beckman, T. J. (2009). Does scale length matter? A comparison of nine-versus five-point rating scales for the mini-CEX. *Advances in health sciences education*, 14(5), 655-664.

²⁴ 3D Group. (2013). Current practices in 360 degree feedback: A benchmark study of North American companies. 3D Group consults regarding a wide range of issues involving 360 degree feedback and employee surveys. 3D Group works with businesses to ensure, among other things, that their 360 degree assessments relate to organizational needs. See <http://www.3dgroup.net/> (last visited May 7, 2014).

Sachs to a materially greater degree than one would expect. I concluded that the distribution of scores is typical, adequate, and useful.

52. Table 1 shows the descriptive statistics on the 360 Feedback ratings at Goldman Sachs for Associates and Vice Presidents by year.²⁵ At least two observations are relevant. First, the difference between the 10% and 90% employees averages more than .80 across years. The difference between the 10% and the 75% employees (top and bottom “quartiles” in the Goldman Sachs rating process) averages about .70. These are meaningful differences, reflecting real perceived differences in job performance. Second, although there was a gradual decrease in variation over time as indicated by the standard deviations²⁶ (from .41 to .25 from 2003 to 2009, the last year in which the 5-point scale was used), the standard deviation increased to .50 when the 9-point scale was introduced in 2010 and remained somewhat higher (.36) in 2011. Thus, increasing the scale increased differentiation, as the firm expected.

53. A close examination of the distributions of actual individual ratings provided in the 360 Feedback process shows a meaningful amount of variation that distinguishes well between levels of job performance. Table 2 shows a frequency distribution of the number and percentage of employees at each level of the rating scales.²⁷ The data from 2009 and 2010 are presented to show the last year the 5-point scale was used and the first year the 9-point scale was used. As can be seen, the 5-point scale effectively distinguishes between two levels of performance, with about 40% receiving a 4 and about 50% receiving a 5. The remaining 7-8% received a 3. Although raters were clearly reluctant to use the entire scale, those raters did make

²⁵ Table 1 was provided by Dr. Michael P. Ward through the attorneys for the defendant.

²⁶ The standard deviation is the commonly used index of the amount of variation in a dataset. It reflects the average difference between individual ratings and the mean rating.

²⁷ Table 2 was condensed from a table provided by Dr. Michael P. Ward through the attorneys for the defendant.

meaningful distinctions. Not everyone got the same ratings, and the feedback derived from the exercise provided meaningful data on employee performance.

54. The 9-point scale adopted in 2010 does an even better job differentiating among levels of performance even though the 5-point scale was not too compressed to be useful (as shown above). Now, four levels of performance are distinguished, with about a third receiving a 9, about a third receiving an 8, about a sixth receiving a 7, and about a sixth receiving a 6. This is a significant spread in an absolute sense and an improvement over the 5-point scale.

55. Many common scales familiar to most people do not use the entire range, yet provide useful (if imperfect) levels of differentiation. For example, grades in colleges and universities theoretically range from A to F (actually A-B-C-D-F, or 5 points), but as a practical matter, most students receive an A or a B, with some Cs. Colleges, universities and employers do not view these ratings as useless or arbitrary. Most people would view the accomplishments of A students as better than those of B students, and both would be viewed as more successful than C students. Using this information, universities and employers routinely identify the most and least accomplished students, even if the rating scale is “compressed.”

56. The variance in Goldman Sachs’ 360 Feedback scores is similar to that reported in the literature. In my Initial Report, I described the results of an extensive literature review on the subject of 360 evaluations (*see* Table 1 to Initial Report).²⁸ Many of the cited articles provide the standard deviations for mean scores on 360 instruments collected from employees. The average standard deviations of the 360 Feedback process ratings at Goldman Sachs are slightly less than, but comparable to, the average observed in 360 review ratings in the literature (.46 in the literature versus .31 for the 5-point scale at Goldman Sachs shown in Table 3 and .44 for the

²⁸ This Table was submitted with the Initial Report on December 11, 2013.

9-point scale in Table 3).²⁹ They are well within the range of variations reported in the literature. In this instance, one would expect the standard deviation to be somewhat smaller because Goldman Sachs uses the 360 Feedback process ratings as one input into employment decisions, rather than merely as a feedback tool. When raters believe their critical input may have an impact on pay or promotion, they are more reluctant to be candid. Thus, ratings used in part for this sort of decision-making tend to reflect smaller standard deviations due partly to the leniency bias described above, as shown in the research.³⁰

57. Finally, Dr. Cascio's new opinion that Goldman Sachs acted "arbitrarily" when it used the 360 Feedback scores in the compensation process seems to rest on a misunderstanding of the role the scores themselves actually play in that process. First, the 360 review scores and feedback are only one input into the manager "quartile" assignment, which in turn is one input into compensation recommendations. In my view, Goldman Sachs does not act arbitrarily in taking these data into account as a part of a multi-factored decision-making process. Second, even as to the 360 Feedback results viewed in isolation, Dr. Cascio appears to have ignored the fact that the scores do not exist in a vacuum; rather, the raters in Goldman Sachs' 360 Feedback process are also asked to provide narrative evaluative comments about the subject employee's performance. Thus, when the manager considers the 360 Feedback data as one input to inform his or her quartile decisions, he or she is not left entirely to the numeric scores; the written comments from reviewers are also considered alongside the ratings. These comments further assist managers in making distinctions among employees who may have closely clustered scores.

²⁹ All citations listed in Table 3 were either included in Table 1 submitted with the Initial Report on December 11, 2013, or are included in the list of References attached to this Report.

³⁰ For example, see Greguras, G. J., Robie, C., Schleicher, D. J., & Maynard III, G. O. F. F. (2003). A field study of the effects of rating purpose on the quality of multisource ratings. *Personnel Psychology*, 56(1), 1-21.

SIGNATURE

A handwritten signature in black ink that reads "M. A. Campion". The signature is written in a cursive, slightly slanted style.

Michael A. Campion, Ph.D.

May 7, 2014

Date

Table 1: Descriptive Statistics of 360 Ratings at Goldman Sachs

Position Title	Fiscal Year	Employees	Mean	Standard Deviation	Minimum	5th Percentile	10th Percentile	25th Percentile	50th Percentile	75th Percentile	90th Percentile	95th Percentile	Maximum
All	2003	614	3.93	0.41	2.33	3.17	3.41	3.69	4.00	4.21	4.41	4.53	5.00
Associate	2003	170	3.92	0.41	2.73	3.10	3.39	3.65	4.00	4.20	4.42	4.50	4.70
VicePresident	2003	444	3.94	0.41	2.33	3.19	3.43	3.70	4.00	4.21	4.41	4.54	5.00
All	2004	1,568	3.92	0.37	1.93	3.29	3.44	3.68	3.93	4.17	4.39	4.49	4.97
Associate	2004	558	3.78	0.38	1.93	3.12	3.32	3.53	3.79	4.05	4.26	4.39	4.97
VicePresident	2004	1,010	3.99	0.35	2.30	3.42	3.58	3.76	3.99	4.24	4.43	4.54	4.91
All	2005	2,438	3.98	0.38	1.90	3.30	3.48	3.75	4.00	4.24	4.44	4.54	5.00
Associate	2005	897	3.84	0.38	2.55	3.18	3.33	3.60	3.86	4.10	4.31	4.45	4.84
VicePresident	2005	1,541	4.06	0.35	1.90	3.45	3.61	3.85	4.08	4.30	4.49	4.59	5.00
All	2006	2,854	4.10	0.28	2.67	3.62	3.74	3.94	4.13	4.29	4.43	4.51	4.88
Associate	2006	1,070	4.01	0.30	2.67	3.51	3.63	3.81	4.03	4.22	4.37	4.46	4.86
VicePresident	2006	1,784	4.16	0.25	2.68	3.73	3.85	4.02	4.18	4.32	4.46	4.53	4.85
All	2007	3,152	4.18	0.27	2.43	3.71	3.85	4.02	4.20	4.37	4.50	4.59	5.00
Associate	2007	1,177	4.12	0.30	2.43	3.61	3.76	3.94	4.13	4.33	4.48	4.58	5.00
VicePresident	2007	1,975	4.22	0.25	2.53	3.80	3.91	4.07	4.23	4.38	4.51	4.59	4.87
All	2008	2,829	4.31	0.24	3.01	3.91	4.01	4.16	4.33	4.48	4.60	4.68	5.00
Associate	2008	941	4.29	0.26	3.25	3.83	3.96	4.12	4.31	4.48	4.60	4.67	4.91
VicePresident	2008	1,888	4.33	0.23	3.01	3.94	4.03	4.18	4.34	4.49	4.60	4.68	5.00
All	2009	2,643	4.43	0.25	2.21	3.97	4.10	4.29	4.46	4.60	4.71	4.78	5.00
Associate	2009	829	4.40	0.27	3.23	3.91	4.04	4.23	4.44	4.59	4.71	4.78	5.00
VicePresident	2009	1,814	4.44	0.24	2.21	4.01	4.14	4.31	4.46	4.60	4.71	4.78	5.00
All	2010	2,737	7.75	0.50	5.29	6.82	7.09	7.46	7.82	8.13	8.32	8.42	8.87
Associate	2010	821	7.66	0.57	5.29	6.57	6.89	7.34	7.77	8.08	8.28	8.39	8.83

VicePresident	2010	1,916	7.79	0.46	5.67	6.96	7.17	7.50	7.84	8.15	8.34	8.43	8.87
All	2011	2,696	8.13	0.36	5.91	7.47	7.66	7.93	8.19	8.38	8.53	8.61	8.95
Associate	2011	840	8.07	0.42	5.91	7.29	7.51	7.85	8.16	8.37	8.52	8.60	8.91
VicePresident	2011	1,856	8.16	0.33	6.50	7.57	7.72	7.97	8.20	8.39	8.53	8.61	8.95

Table 2: Frequency Distributions of 360 Ratings

2009						
Score	All		Associate		Vice President	
	N	%	N	%	N	%
1	307	0.1%	57	0.1%	250	0.1%
2	1,475	0.4%	432	0.4%	1,043	0.4%
3	28,096	7.5%	8,279	8.1%	19,817	7.3%
4	153,772	41.1%	41,728	40.8%	112,044	41.2%
5	190,272	50.9%	51,780	50.6%	138,492	51.0%
Total	373,922	-	102,276	-	271,646	-

2010						
Score	All		Associate		Vice President	
	N	%	N	%	N	%
1	58	0.0%	22	0.0%	36	0.0%
2	149	0.0%	66	0.1%	83	0.0%
3	614	0.2%	250	0.3%	364	0.1%
4	2,530	0.7%	883	0.9%	1,647	0.6%
5	10,254	2.8%	3,118	3.3%	7,136	2.6%
6	56,422	15.4%	15,095	15.9%	41,327	15.3%
7	61,037	16.7%	15,922	16.8%	45,115	16.7%
8	124,665	34.1%	31,551	33.3%	93,114	34.4%
9	109,616	30.0%	27,900	29.4%	81,716	30.2%
Total	365,345	-	94,807	-	270,538	-

Note: Data are aggregated across items of the 360 instruments and across employees.

Table 3: Standard Deviations of 360 Ratings in the Literature

Citation*	Standard Deviation (Mean Scores on 5-Point Scales or as Noted; Excludes self-Ratings)
Antonioni & Park (2001)	.51 (20 items)
Antonioni (1995)	Average of .47 (7 items)
Atkins & Wood (2002)	.35 (46 items)
Atwater & Brett (2006)	.50 (47 items, 9-point)
Atwater & Yammarino (1993)	Average of .57 (24 items)
Atwater et al. (1998)	Average of .40 (106 items)
Atwater et al. (2000)	Average of .67 (43 items)
Atwater et al. (2007)	Average of .45 (196 items)
Atwater et al. (2009)	Average of .39 (196 items)
Bailey & Austin (2006)	Average of .43 (55 items)
Bailey & Fletcher (2002)	Average of .43 (50 items, 6-point)
Bono & Colbert (2005)	.34 (16 items)
Church (2000)	Average of .60 (61 items, 7-point)
Church, Rogelberg, & Wacławski (2000)	Average of .44 (9 items)
Eckert et al (2010)	Average of .48 (4-11 items)
Fleenor, McCauley, & Brutus (1996)	Average of .44 (106 items)
Fletcher & Baldry (2000)	.42 (36 items, 6-point)
Furnham & Stringfield (1998)	Average of .79 (4-8 items, 7-point)
Gentry, Yip, & Hannum (2010)	Average of .48 (40 items)
Goffin & Anderson (2007)	Average of .58 (88 items, 7-point)
Harris, Smith, & Champagne (1995)	Average of .69 (9 items)
Heslin & Latham (2004)	Average of .46 (32 items)
Johnson & Ferstl (1999)	Average of .36 (36 items)
Maurer, Mitchell, & Barbeite (2002)	Average of .36 (27 items)
Ostroff, Atwater, & Feinberg (2004)	Average of .43 (85 items)
Reilly Smither & Vasilopoulos (1996)	Average of .58 (33 items)
Seifert & Yuki (2010)	Average of .51 (4 scales)

Smither et.al (2003)	Average of .39 (8 items)
vanHooff, Flier, & Minne (2006)	Average of .41 (14 items)
Walker & Smither (1999)	Average of .44 (29 items)
Wohlens et al. (1993)	Average of .49 (47 items)
Average (All)	.48
Average (5-point scales only)	.46
Average at Goldman Sachs (5-point scale)	.31
Average at Goldman Sachs (9-point scale)	.44

*For full citations, see Table 1 in Initial Report and the list of References attached to this Report.

General References in Sur-Rebuttal Report

Barrett, G. V., & Kernan, M. C. (1987). Performance appraisal and terminations: A review of court decisions since *Brito v. Zia* with implications for personnel practices. *Personnel Psychology*, 40(3), 489-503.

Bono, J. E., & Colbert, A. E. (2005). Understanding Responses to Multi-Source Feedback: The Role of Self-Evaluations. *Personnel Psychology*, 58(1), 171-203.

Campion, M. A. (1993). Article review checklist: A criterion checklist for reviewing research articles in applied psychology. *Personnel Psychology*, 46, 705-718.

Cascio, W. F., & Valenzi, E. R. (1977). Behaviorally anchored rating scales: Effects of education and job experience of raters and ratees. *Journal of Applied Psychology*, 62(3), 278.

Cascio, W. F., & Aguinis, H. (2005). *Applied psychology in human resource management* (6th ed.). Upper Saddle River, NJ: Pearson Prentice-Hall (see p. 97).

Cascio, W. F., & Aguinis, H. (2008). Research in industrial and organizational psychology from 1963 to 2007: changes, choices, and trends. *Journal of Applied Psychology*, 93(5), 1062-1081.

Cook, D. A., & Beckman, T. J. (2009). Does scale length matter? A comparison of nine-versus five-point rating scales for the mini-CEX. *Advances in health sciences education*, 14(5), 655-664.

Eckert, R., Ekelund, B. Z., Gentry, W. A., & Dawson, J. F. (2010). "I don't see me like you see me, but is that a problem?" Cultural influences on rating discrepancy in 360-degree feedback instruments. *European Journal of Work and Organizational Psychology*, 19(3), 259-278.

Faley, R.H., Kleiman, L.S., & Lengnick-Hall, M.L. (1984). Age discrimination and personnel psychology: A review and synthesis of the legal literature with implications of future research. *Personnel Psychology*, 37, 327-350.

Field, H.S., & Holley, W.H. (1982). The relationship of performance appraisal system characteristics to verdicts in selected employment discrimination cases. *Academy of Management Journal*, 25, 392-406.

Greguras, G. J., Robie, C., Schleicher, D. J., & Maynard III, G. O. F. F. (2003). A field study of the effects of rating purpose on the quality of multisource ratings. *Personnel Psychology*, 56(1), 1-21.

Hollenbeck, J. R., Gerhart, B., & Wright, P. M. (2004). *Fundamentals of human resource management* (4th ed). New York: McGraw-Hill Irwin.).

Kleiman, R.L., & Durham R.L. (1981).). Performance appraisal, promotion and the courts: A critical review. *Personnel Psychology*, 34, 103-121.

Krosnick, J. A., & Fabrigar, L. R. (1997). Designing rating scales for effective measurement in surveys. *Survey measurement and process quality*, 141-164.

Maurer, T. J., Mitchell, D. R., & Barbeite, F. G. (2002). Predictors of attitudes toward a 360-degree feedback system and involvement in post-feedback management development activity. *Journal of Occupational and Organizational Psychology*, 75(1), 87-107.

Preston, C. C., & Colman, A. M. (2000). Optimal number of response categories in rating scales: reliability, validity, discriminating power, and respondent preferences. *Acta psychologica*, 104(1), 1-15.

References Citing the Principles for the Validation and Use of Personnel Selection Procedures

Arnold, J. D., Rauschenberger, J. M., Soubel, W. G., & Guion, R. M. (1982). Validation and utility of a strength test for selecting steelworkers. *Journal Of Applied Psychology*, 67(5), 588-604.

Barrett, G. V., Phillips, J. S., & Alexander, R. A. (1981). Concurrent and predictive validity designs: A critical reanalysis. *Journal Of Applied Psychology*, 66(1), 1-6.

Benjamin, L. r. (1997). A history of Division 14 (The Society for Industrial and Organizational Psychology). In D. A. Dewsbury (Ed.) *Unification through division: Histories of the divisions of the American Psychological Association*, Vol. 2 (pp. 101-126). Washington, DC US: American Psychological Association.

Bing, M. N., Davison, H., & Arvey, R. D. (2009). Using a repeated-measures approach to validating personality tests in small samples: A feasibility study with implications for small businesses. *Journal Of Managerial Issues*, 21(1), 11-35.

Boehm, V. R. (1977). Differential prediction: A methodological artifact? *Journal Of Applied Psychology*, 62(2), 146-154.

Bryan, L., & Vinchur, A. J. (2013). Industrial-organizational psychology. In D. K. Freedheim, I. B. Weiner (Eds.), *Handbook of psychology*, Vol. 1: History of psychology (2nd ed.) (pp. 407-428). Hoboken, NJ US: John Wiley & Sons Inc.

Burke, M. J., & Normand, J. (1987). Computerized psychological testing: Overview and critique. *Professional Psychology: Research And Practice*, 18(1), 42-51.

Casebook for providers of psychological services. (1982). *American Psychologist*, 37(6), 698-701.

Cronshaw, S. F. (1986). The status of employment testing in Canada: A review and evaluation of theory and professional practice. *Canadian Psychology/Psychologie Canadienne*, 27(2), 183-195.

Culpepper, S. (2012). Using the criterion-predictor factor model to compute the probability of detecting prediction bias with ordinary least squares regression. *Psychometrika*, 77(3), 561-580.

Eyde, L. D., Robertson, G. J., & Krug, S. E. (2010). *Responsible test use: Case studies for assessing human behavior* (2nd ed.). Washington, DC US: American Psychological Association.

Gullickson, A. R., & Howard, B. B. (2009). *The personnel evaluation standards: How to assess systems for evaluating educators* (2nd ed.). Thousand Oaks, CA US: Corwin Press.

Hanges, P. J., Salmon, E. D., & Aiken, J. R. (2013). Legal issues in industrial testing and assessment. In K. F. Geisinger, B. A. Bracken, J. F. Carlson, J. C. Hansen, N. R. Kuncel, S. P. Reise, M. C. Rodriguez (Eds.) *APA handbook of testing and assessment in psychology*, Vol. 1:

Test theory and testing and assessment in industrial and organizational psychology (pp. 693-711). Washington, DC US: American Psychological Association.

Hogan, J., & Quigley, A. M. (1986). Physical standards for employment and the courts. *American Psychologist*, 41(11), 1193-1217.

Klimoski, R. J., & Wilkinson, T. R. (2013). Psychological assessment in industrial/organizational settings. In J. R. Graham, J. A. Naglieri, I. B. Weiner (Eds.) , *Handbook of psychology*, Vol. 10: Assessment psychology (2nd ed.) (pp. 347-372). Hoboken, NJ US: John Wiley & Sons Inc.

Landy, F. J. (1986). Stamp collecting versus science: Validation as hypothesis testing. *American Psychologist*, 41(11), 1183-1192.

Landy, F. J. (2003). Validity generalization: Then and now. In K. R. Murphy (Ed.) , *Validity generalization: A critical review* (pp. 155-195). Mahwah, NJ US: Lawrence Erlbaum Associates Publishers.

Lawshe, C. H. (1985). Inferences from personnel tests and their validity. *Journal Of Applied Psychology*, 70(1), 237-238.

Lawshe, C. H. (1987). Adverse impact: Is it a viable concept?. *Professional Psychology: Research And Practice*, 18(5), 492-497.

Messick, S. (1980). Test validity and the ethics of assessment. *American Psychologist*, 35(11), 1012-1027.

Mitchell, T. W., & Klimoski, R. J. (1986). Estimating the validity of cross-validity estimation. *Journal Of Applied Psychology*, 71(2), 311-317.

Novick, M. R. (1981). Federal guidelines and professional standards. *American Psychologist*, 36(10), 1035-1046.

Posthuma, R. A. (2002). Employee selection procedures and the business necessity defense. *Applied H.R.M. Research*, 7(1-2), 53-63.

Ree, M., Carretta, T. R., & Steindl, J. R. (2002). Cognitive ability. In N. Anderson, D. S. Ones, H. Sinangil, C. Viswesvaran (Eds.) , *Handbook of industrial, work and organizational psychology*, Volume 1: Personnel psychology (pp. 219-232). Thousand Oaks, CA: Sage Publications Ltd.

Sands, W. A., Waters, B. K., & McBride, J. R. (1997). Computerized adaptive testing: From inquiry to operation. Washington, DC US: American Psychological Association.

Schmidt, F. L. (1992). What do data really mean? Research findings, meta-analysis, and cumulative knowledge in psychology. *American Psychologist*, 47(10), 1173-1181.

Schmidt, F. L., & Hunter, J. E. (1977). Development of a general solution to the problem of validity generalization. *Journal Of Applied Psychology*, 62(5), 529-540.

Schmidt, F. L., & Hunter, J. E. (1981). Employment testing: Old theories and new research findings. *American Psychologist*, 36(10), 1128-1137.

Schmidt, F. L., Hunter, J. E., & Outerbridge, A. N. (1986). Impact of job experience and ability on job knowledge, work sample performance, and supervisory ratings of job performance. *Journal Of Applied Psychology*, 71(3), 432-439.

Schmitt, N., Coyle, B. W., & Rauschenberger, J. (1977). A Monte Carlo evaluation of three formula estimates of cross-validated multiple correlation. *Psychological Bulletin*, 84(4), 751-758.

Shimberg, B. (1981). Testing for licensure and certification. *American Psychologist*, 36(10), 1138-1146.

Specialty Guidelines for the Delivery of Services by Industrial/Organizational Psychologists. (1981). *American Psychologist*, 36(6), 664-669.

Stone, E. F., & Hollenbeck, J. R. (1989). Clarifying some controversial issues surrounding statistical procedures for detecting moderator variables: Empirical evidence and related matters. *Journal Of Applied Psychology*, 74(1), 3-10.

Stone-Romero, E. F., & Anderson, L. E. (1994). Relative power of moderated multiple regression and the comparison of subgroup correlation coefficients for detecting moderating effects. *Journal Of Applied Psychology*, 79(3), 354-359.

Stone-Romero, E. F., & Liakhovitski, D. (2002). Strategies for Detecting Moderator Variables: A Review of Conceptual and Empirical Issues. In G. R. Ferris, J. J. Martocchio (Eds.) , *Research in personnel and human resources management* (pp. 333-372). US: Elsevier Science/JAI Press.

Tenopyr, M. L. (1981). The realities of employment testing. *American Psychologist*, 36(10), 1120-1127.

Vinchur, A. J. (2007). A History of Psychology Applied to Employee Selection. In L. L. Koppes (Ed.) , *Historical perspectives in industrial and organizational psychology* (pp. 193-218). Mahwah, NJ US: Lawrence Erlbaum Associates Publishers.

References Citing the Uniform Guidelines on Employee Selection Procedures

Aguinis, H., & Smith, M. A. (2010). Balancing adverse impact, selection errors, and employee performance in the presence of test bias. In J. L. Outtz (Ed.) , *Adverse impact: Implications for organizational staffing and high stakes selection* (pp. 403-423). New York, NY US: Routledge/Taylor & Francis Group.

Aiken, J. R., Salmon, E. D., & Hanges, P. J. (2013). The origins and legacy of the Civil Rights Act of 1964. *Journal Of Business And Psychology*, 28(4), 383-399.

Aramburu-Zabala Higuera, L. (2001). Adverse impact in personnel selection: The legal framework and test bias. *European Psychologist*, 6(2), 103-111.

Aramburu-Zabala, L. A. (2004). La Directiva Antidiscriminatoria (2000/78/EC): Implicaciones en selección de personal. *Revista De Psicología Del Trabajo Y De Las Organizaciones*, 20(2), 199-223.

Aramburu-Zabala, L., & Casals, M. (2001). Predictive validity and adverse impact in the professional certification programmes: The case of gender differences. *European Review Of Applied Psychology/Revue Européenne De Psychologie Appliquée*, 51(3), 179-187.

Aramburu-Zabala, L., & Casals, M. (2003). Prediction of training performance for diesel mechanics. *Swiss Journal Of Psychology/Schweizerische Zeitschrift Für Psychologie/Revue Suisse De Psychologie*, 62(4), 233-240.

Arthur, W. r. (2012). Dimension-based assessment centers: Theoretical perspectives. In D. R. Jackson, C. E. Lance, B. J. Hoffman (Eds.) , *The psychology of assessment centers* (pp. 95-120). New York, NY US: Routledge/Taylor & Francis Group.

Arvey, R. D. (1979). Unfair discrimination in the employment interview: Legal and psychological aspects. *Psychological Bulletin*, 86(4), 736-765.

Avolio, B. J., & Waldman, D. A. (1989). Ratings of managerial skill requirements: Comparison of age- and job-related factors. *Psychology And Aging*, 4(4), 464-470.

Banks, G. C., & McDaniel, M. A. (2012). Meta-analysis as a validity summary tool. In N. Schmitt (Ed.) , *The Oxford handbook of personnel assessment and selection* (pp. 156-175). New York, NY US: Oxford University Press.

Baranowski, L. E., & Anderson, L. E. (2005). Examining rating source variation in work behavior to KSA linkages. *Personnel Psychology*, 58(4), 1041-1054.

Barrett, G. V. (2000). Personnel selection: Selection and the law. In A. E. Kazdin (Ed.) , *Encyclopedia of psychology*, Vol. 6 (pp. 156-160). Washington, DC New York, NY USUS: American Psychological Association.

Barrett, G. V., Miguel, R. F., & Doverspike, D. (2011). The Uniform Guidelines: Better the devil you know. *Industrial And Organizational Psychology: Perspectives On Science And Practice*, 4(4), 534-536.

Barrett, K. M. (1996). Tips for assuring the validity of inferences based on job analysis procedures. *International Journal Of Selection And Assessment*, 4(2), 87-95.

Barrett, K. M. (1996). Tips for assuring the validity of inferences based on job analysis procedures. *International Journal Of Selection And Assessment*, 4(2), 87-95.

Bartram, D. (1995). Predicting adverse impact in selection testing. *International Journal Of Selection And Assessment*, 3(1), 52-61.

Bernardin, H., Beatty, R. W., & Jensen, W. (1980). The new Uniform Guidelines on Employee Selection Procedures in the context of university personnel decisions. *Personnel Psychology*, 33(2), 301-316.

Biddle, D. A. (2010). Should employers rely on local validation studies or validity generalization (VG) to support the use of employment tests in Title VII situations?. *Public Personnel Management*, 39(4), 307-326.

Binning, J. F., & Barrett, G. V. (1989). Validity of personnel decisions: A conceptual analysis of the inferential and evidential bases. *Journal Of Applied Psychology*, 74(3), 478-494.

Bobko, P., & Roth, P. L. (2010). An analysis of two methods for assessing and indexing adverse impact: A disconnect between the academic literature and some practice. In J. L. Outtz (Ed.) , *Adverse impact: Implications for organizational staffing and high stakes selection* (pp. 29-49). New York, NY US: Routledge/Taylor & Francis Group.

Brannick, M. T., Levine, E. L., & Morgeson, F. P. (2007). *Job and work analysis: Methods, research, and applications for human resource management* (2nd ed.). Thousand Oaks, CA US: Sage Publications, Inc.

Bricklin, P. M., & Ciuccio, J. (2003). Psychopharmacology examination for psychologists developed by the American Psychological Association Practice Organization's College of Professional Psychology. In M. T. Sammons, R. Paige, R. F. Levant (Eds.) , *Prescriptive authority for psychologists: A history and guide* (pp. 179-189). Washington, DC US: American Psychological Association.

Brink, K. E., & Crenshaw, J. L. (2011). The affronting of the Uniform Guidelines: From propaganda to discourse. *Industrial And Organizational Psychology: Perspectives On Science And Practice*, 4(4), 547-553.

Broach, D., Farmer, W. L., & Young, W. C. (1999). Differential prediction of FAA Academy performance on the basis of race and written air traffic control specialist aptitude test scores. *FAA Office Of Aviation Medicine Reports*, DOT-FAA-AM-99-16.

Brown, S. H. (1978). Long-term validity of a personal history item scoring procedure. *Journal Of Applied Psychology*, 63(6), 673-676.

Bryan, L., & Vinchur, A. J. (2012). A history of industrial and organizational psychology. In S. J. Kozlowski (Ed.), *The Oxford handbook of organizational psychology*, Vol. 1 (pp. 22-75). New York, NY US: Oxford University Press.

Buckendahl, C. W., & Hunt, R. (2005). Whose Rules? The Relation Between the 'Rules' and 'Law' of Testing. In R. P. Phelps (Ed.), *Defending standardized testing* (pp. 147-158). Mahwah, NJ US: Lawrence Erlbaum Associates Publishers.

Burns, W. C. (1996). Content validity, face validity, and quantitative face validity. In R. S. Barrett (Ed.) *Fair employment strategies in human resource management* (pp. 38-46). Westport, CT US: Quorum Books/Greenwood Publishing Group.

Buster, M. A., Roth, P. L., & Bobko, P. (2005). A Process For Content Validation Of Education And Experienced-Based Minimum Qualifications: An Approach Resulting In Federal Court Approval. *Personnel Psychology*, 58(3), 771-799.

Butcher, J. N., Gucker, D. K., & Hellervik, L. W. (2009). Clinical personality assessment in the employment context. In J. N. Butcher (Ed.), *Oxford handbook of personality assessment* (pp. 582-598). New York, NY US: Oxford University Press.

Camara, W. J. (1992). Fairness and fair use in employment testing: A matter of perspective. In K. F. Geisinger (Ed.) *Psychological testing of Hispanics* (pp. 215-231). Washington, DC US: American Psychological Association.

Camara, W. J. (1996). Fairness and public policy in employment testing: Influences from a professional association. In R. S. Barrett (Ed.), *Fair employment strategies in human resource management* (pp. 3-11). Westport, CT US: Quorum Books/Greenwood Publishing Group.

Camara, W. J., & Merenda, P. F. (2000). Using personality tests in preemployment screening: Issues raised in *Soroka v. Dayton Hudson Corporation*. *Psychology, Public Policy, And Law*, 6(4), 1164-1186.

Camilli, G. (2013). Ongoing issues in test fairness. *Educational Research And Evaluation*, 19(2-3), 104-120.

Campbell, J. P., & Knapp, D. J. (2001). Exploring the limits in personnel selection and classification. Mahwah, NJ US: Lawrence Erlbaum Associates Publishers.

Carnahan, T. J. (2007). A methodological look at structured interview reliability: Interrater agreement versus parallel forms. *Dissertation Abstracts International*, 68.

Carson, J. (2001). Legal issues in standard setting for licensure and certification. In G. J. Cizek (Ed.), *Setting performance standards: Concepts, methods, and perspectives* (pp. 427-444). Mahwah, NJ US: Lawrence Erlbaum Associates Publishers.

Carter, G. W., Cook, K. W., & Dorsey, D. W. (2009). *Career paths: Charting courses to success for organizations and their employees*. Wiley-Blackwell.

Cattell, H. P., & Schuerger, J. M. (2003). *Essentials of 16PF assessment*. Hoboken, NJ US: John Wiley & Sons Inc.

Collins, W. E., Boone, J. O., & VanDeventer, A. D. (1981). The selection of air traffic control specialists: History and review of contributions by the Civil Aeromedical Institute, 1960–80. *Aviation, Space, And Environmental Medicine*, 52(4), 217-240.

Cuttler, M. J. (2011). Pre-employment screening of police officers: Integrating actuarial prediction models with practice. In J. Kitaeff (Ed.) *Handbook of police psychology* (pp. 135-163). New York, NY US: Routledge/Taylor & Francis Group.

De Soete, B., Lievens, F., Oostrom, J., & Westerveld, L. (2013). Alternative predictors for dealing with the diversity–validity dilemma in personnel selection: The constructed response multimedia test. *International Journal Of Selection And Assessment*, 21(3), 239-250.

Dean, M. A. (2013). Examination of ethnic group differential responding on a biodata instrument. *Journal Of Applied Social Psychology*, 43(9), 1905-1917.

Dipboye, R. L., & Johnson, S. K. (2013). Understanding and improving employee selection interviews. In K. F. Geisinger, B. A. Bracken, J. F. Carlson, J. C. Hansen, N. R. Kuncel, S. P. Reise, M. C. Rodriguez (Eds.) *APA handbook of testing and assessment in psychology*, Vol. 1: *Test theory and testing and assessment in industrial and organizational psychology* (pp. 479-499). Washington, DC US: American Psychological Association.

Doverspike, D., & Arthur, W.R. (2012). The role of job analysis in test selection and development. In M. A. Wilson, W. r. Bennett, S. G. Gibson, G. M. Alliger (Eds.) , *The handbook of work analysis: Methods, systems, applications and science of work measurement in organizations* (pp. 381-399). New York, NY US: Routledge/Taylor & Francis Group.

Doverspike, D., Taylor, M., & Arthur, W. r. (2006). *Psychological perspective on affirmative action*. New York, NY US: Novinka/Nova Science Publishers.

Dunleavy, E. M., Aamodt, M. G., Morgan, D. A., Gutman, A., & Cohen, D. B. (2011). Guidelines, principles, standards, and the courts: Why can't they all just get along?. *Industrial And Organizational Psychology: Perspectives On Science And Practice*, 4(4), 515-520.

Dunleavy, E. M., Mueller, L. M., Buonasera, A. K., Kuang, D. C., & Dunleavy, D. (2008). On the consequences of frequent applicants in adverse impact analyses: A demonstration study. *International Journal Of Selection And Assessment*, 16(4), 333-344.

Dunleavy, E. M., Stuebing, K. M., Campion, J. E., & Glenn, D. M. (2008). Using the 4/5ths rule as an outcome in regression analyses: A demonstrative simulation. *Journal Of Business And Psychology*, 23(3-4), 103-114.

Earles, J. A., Driskill, W. E., & Dittmar, M. J. (1996). Methodology for identifying abilities for job classification: An application of job analysis. *Military Psychology*, 8(3), 179-193.

Epstein, S. (1995). What can be done to improve the journal review process. *American Psychologist*, 50(10), 883-885.

Fernández Garrido, J., & Aramburu-Zabala Higuera, L. (2000). Diferencias de género en los procesos de selección para la formación para el Empleo. Un estudio empírico. *Revista De Psicología Del Trabajo Y De Las Organizaciones*, 16(3), 329-340.

Foster, J., Gaddis, B., & Hogan, J. (2012). Personality-based job analysis. In M. A. Wilson, W. r. Bennett, S. G. Gibson, G. M. Alliger (Eds.) *The handbook of work analysis: Methods, systems, applications and science of work measurement in organizations* (pp. 247-264). New York, NY US: Routledge/Taylor & Francis Group.

Garman, A. N. (2002). Assessing candidates for leadership positions. In R. L. Lowman (Ed.) , *The California School of Organizational Studies: Handbook of organizational consulting psychology: A comprehensive guide to theory, skills, and techniques* (pp. 185-211). San Francisco, CA US: Jossey-Bass.

Gebhardt, D. L., & Baker, T. A. (2010). Physical performance tests. In J. L. Farr, N. T. Tippins (Eds.) , *Handbook of employee selection* (pp. 277-298). New York, NY US: Routledge/Taylor & Francis Group.

Geisinger, K. F., & McCormick, C. M. (2010). Adopting cut scores: Post-standard-setting panel considerations for decision makers. *Educational Measurement: Issues And Practice*, 29(1), 38-44.

Gelerter, R. (1989). The uniform guidelines and subjective selection criteria and procedures: Conference remarks. In B. R. Gifford (Ed.) , *Test policy and the politics of opportunity allocation: The workplace and the law* (pp. 127-132). New York, NY US: Kluwer Academic/Plenum Publishers.

Gibson, W. M., & Caplinger, J. A. (2007). Transportation of Validation Results. In S. McPhail (Ed.) , *Alternative validation strategies: Developing new and leveraging existing validity evidence* (pp. 29-81). Hoboken, NJ US: John Wiley & Sons Inc.

Goldstein, B. L., & Patterson, P. O. (1988). Turning back the Title VII clock: The resegregation of the American work force through validity generalization. *Journal Of Vocational Behavior*, 33(3), 452-462.

Guion, R. M. (1980). On Trinitarian doctrines of validity. *Professional Psychology*, 11(3), 385-398.

Guion, R. M., & Highhouse, S. (2006). *Essentials of personnel assessment and selection*. Mahwah, NJ US: Lawrence Erlbaum Associates Publishers.

Hamdani, M., Valcea, S., & Buckley, M. (2014). The relentless pursuit of construct validity in the design of employment interviews. *Human Resource Management Review*, 24(2), 160-176.

Hanges, P. J., Aiken, J. R., & Salmon, E. D. (2011). The devil is in the details (and the context): A call for care in discussing the Uniform Guidelines. *Industrial And Organizational Psychology: Perspectives On Science And Practice*, 4(4), 562-565.

Hanges, P. J., Salmon, E. D., & Aiken, J. R. (2013). Legal issues in industrial testing and assessment. In K. F. Geisinger, B. A. Bracken, J. F. Carlson, J. C. Hansen, N. R. Kuncel, S. P. Reise, M. C. Rodriguez (Eds.) *APA handbook of testing and assessment in psychology*, Vol. 1: Test theory and testing and assessment in industrial and organizational psychology (pp. 693-711). Washington, DC US: American Psychological Association.

Harris, E. G., Artis, A. B., Fogliasso, C., & Fleming, D. E. (2006). Hospital employee job resourcefulness: An empirical study and implications for health care marketing. *Health Marketing Quarterly*, 24(1-2), 63-75.

Highhouse, S. (2009). Tests don't measure jobs: The meaning of content validation. *Industrial And Organizational Psychology: Perspectives On Science And Practice*, 2(4), 493-495.

Hoffman, C. C., Tashima, C., & Luck, G. (2010). Using a difficulty-anchored rating scale in performing Angoff ratings. *International Journal Of Selection And Assessment*, 18(4), 407-416.

Horvath, M. (2013). Spiritual gifts inventories: A psychometric perspective. *Journal Of Psychology And Christianity*, 32(2), 124-133.

Inwald, R. E. (1985). Administrative, legal, and ethical practices in the psychological testing of law enforcement officers. *Journal Of Criminal Justice*, 13(4), 367-372.

Jacob, R., Deckert, P. J., & Silva, J. (2011). Adverse impact is far more complicated than the Uniform Guidelines indicate. *Industrial And Organizational Psychology: Perspectives On Science And Practice*, 4(4), 558-561.

Jacobs, R., Murphy, K., & Silva, J. (2013). Unintended consequences of EEO enforcement policies: Being big is worse than being bad. *Journal Of Business And Psychology*, 28(4), 467-471.

Jeanneret, P., & Zedeck, S. (2010). Professional guidelines/standards. In J. L. Farr, N. T. Tippins (Eds.) , *Handbook of employee selection* (pp. 593-625). New York, NY US: Routledge/Taylor & Francis Group.

Johnson, J. W., Steel, P., Scherbaum, C. A., Hoffman, C. C., Jeanneret, P., & Foster, J. (2010). Validation is like motor oil: Synthetic is better. *Industrial And Organizational Psychology: Perspectives On Science And Practice*, 3(3), 305-328.

Joseph, D. L., & Newman, D. A. (2010). Emotional intelligence: An integrative meta-analysis and cascading model. *Journal Of Applied Psychology*, 95(1), 54-78.

Kehoe, J. F. (2010). Cut scores and adverse impact. In J. L. Outtz (Ed.) , *Adverse impact: Implications for organizational staffing and high stakes selection* (pp. 289-322). New York, NY US: Routledge/Taylor & Francis Group.

Kleiman, L. S., & Faley, R. H. (1985). The implications of professional and legal guidelines for court decisions involving criterion-related validity: A review and analysis. *Personnel Psychology*, 38(4), 803-833.

Klimoski, R. J., & Wilkinson, T. R. (2013). Psychological assessment in industrial/organizational settings. In J. R. Graham, J. A. Naglieri, I. B. Weiner (Eds.) , *Handbook of psychology*, Vol. 10: Assessment psychology (2nd ed.) (pp. 347-372). Hoboken, NJ US: John Wiley & Sons Inc.

Kravitz, D. A. (2008). The diversity-validity dilemma: Beyond selection--The role of affirmative action. *Personnel Psychology*, 61(1), 173-193.

Kuthy, J. E. (2009). Reducing adverse impact: An investigation of the effect of additional study time on trainability test performance. *Dissertation Abstracts International*, 69.

Latham, G. P., & Sue-Chan, C. (1996). A legally defensible interview for selecting the best. In R. S. Barrett (Ed.) , *Fair employment strategies in human resource management* (pp. 134-143). Westport, CT US: Quorum Books/Greenwood Publishing Group.

Lawshe, C. H. (1987). Adverse impact: Is it a viable concept?. *Professional Psychology: Research And Practice*, 18(5), 492-497.

Lefkowitz, J. (2003). *Ethics and values in industrial-organizational psychology*. Mahwah, NJ US: Lawrence Erlbaum Associates Publishers.

Lefkowitz, J. (2012). Ethics in industrial–organizational psychology. In S. J. Knapp, M. C. Gottlieb, M. M. Handelsman, L. D. VandeCreek (Eds.) , *APA handbook of ethics in psychology*, Vol 2: Practice, teaching, and research (pp. 149-167). Washington, DC US: American Psychological Association.

Lefkowitz, J., & Lowman, R. L. (2010). Ethics of employee selection. In J. L. Farr, N. T. Tippins (Eds.) , *Handbook of employee selection* (pp. 571-591). New York, NY US: Routledge/Taylor & Francis Group.

Licata, J. W., Mowen, J. C., Harris, E. G., & Brown, T. J. (2003). On the Trait Antecedents and Outcomes of Service Worker Job Resourcefulness: A Hierarchical Model Approach. *Journal Of The Academy Of Marketing Science*, 31(3), 256-271.

Malinowski, F. A. (1980). Test passing points in state and municipal agencies. *Public Personnel Management*, 9(4), 274-277.

Marvin, R. S., & Schutzy, B. M. (2009). One component of an evidence-based approach to the use of attachment research in child custody evaluations. *Journal Of Child Custody: Research, Issues, And Practices*, 6(1-2), 113-138.

Maxwell, S. E., & Arvey, R. D. (1993). The search for predictors with high validity and low adverse impact: Compatible or incompatible goals?. *Journal Of Applied Psychology*, 78(3), 433-437.

McDaniel, M. A. (2007). Validity Generalization as a Test Validation Approach. In S. McPhail (Ed.) , *Alternative validation strategies: Developing new and leveraging existing validity evidence* (pp. 159-180). Hoboken, NJ US: John Wiley & Sons Inc.

McDaniel, M. A. (2009). Gerrymandering in personnel selection: A review of practice. *Human Resource Management Review*, 19(3), 263-270.

McDaniel, M. A., Kepes, S., & Banks, G. C. (2011). Encouraging debate on the Uniform Guidelines and the disparate impact theory of discrimination. *Industrial And Organizational Psychology: Perspectives On Science And Practice*, 4(4), 566-570.

McDaniel, M. A., Kepes, S., & Banks, G. C. (2011). The Uniform Guidelines are a detriment to the field of personnel selection. *Industrial And Organizational Psychology: Perspectives On Science And Practice*, 4(4), 494-514.

McDaniel, M. A., Psotka, J., Legree, P. J., Yost, A., & Weekley, J. A. (2011). Toward an understanding of situational judgment item validity and group differences. *Journal Of Applied Psychology*, 96(2), 327-336.

McGonigle, T. P., & Curnow, C. K. (2002). Development of a modified improved point method experience questionnaire. *Applied H.R.M. Research*, 7(1-2), 15-21.

McPhail, S. (2005). Auditing Selection Processes: Application of a Risk Assessment Model. *The Psychologist-Manager Journal*, 8(2), 205-221.

McPhail, S. (2007). *Alternative validation strategies: Developing new and leveraging existing validity evidence*. Hoboken, NJ US: John Wiley & Sons Inc.

Mead, A. D., & Morris, S. B. (2011). About babies and bathwater: Retaining core principles of the Uniform Guidelines. *Industrial And Organizational Psychology: Perspectives On Science And Practice*, 4(4), 554-557.

Messick, S. (1980). Test validity and the ethics of assessment. *American Psychologist*, 35(11), 1012-1027.

Miller, B. K., Nicols, K., & Eure, J. (2009). Body art in the workplace: Piercing the prejudice?. *Personnel Review*, 38(6), 621-640.

Morgeson, F. P., Spitzmuller, M., Garza, A. S., & Campion, M. A. (2014). Pay attention! The liabilities of respondent experience and carelessness when making job analysis judgments. *Journal Of Management*.

Murphy, K. R., & Jacobs, R. R. (2012). Using effect size measures to reform the determination of adverse impact in equal employment litigation. *Psychology, Public Policy, And Law*.

- Newman, D. A., Jacobs, R. R., & Bartram, D. (2007). Choosing the best method for local validity estimation: Relative accuracy of meta-analysis versus a local study versus Bayes-analysis. *Journal Of Applied Psychology*, 92(5), 1394-1413.
- Norton, S. D., & Gustafson, D. P. (1982). Industrial/organizational psychology as applied to human resources management. *Professional Psychology*, 13(6), 904-917.
- Norton, S. D., Balloun, J. L., & Konstantinovich, B. (1980). The soundness of supervisory ratings as predictors of managerial success. *Personnel Psychology*, 33(2), 377-388.
- Novick, M. R. (1981). Federal guidelines and professional standards. *American Psychologist*, 36(10), 1035-1046.
- O'Boyle, E. r., & McDaniel, M. A. (2009). Criticisms of employment testing: A commentary. In R. P. Phelps (Ed.) , *Correcting fallacies about educational and psychological testing* (pp. 181-197). Washington, DC US: American Psychological Association.
- Outtz, J. L. (2011). Abolishing the Uniform Guidelines: Be careful what you wish for. *Industrial And Organizational Psychology: Perspectives On Science And Practice*, 4(4), 526-533.
- Patterson, P. O. (1989). Employment testing and Title VII of the Civil Rights Act of 1964. In B. R. Gifford (Ed.) , *Test policy and the politics of opportunity allocation: The workplace and the law* (pp. 83-120). New York, NY US: Kluwer Academic/Plenum Publishers.
- Ployhart, R. E. (2012). The content validity of cognitively oriented tests: Commentary on Schmidt (2012). *International Journal Of Selection And Assessment*, 20(1), 19-23.
- Ployhart, R. E., & Ehrhart, M. G. (2002). Modeling the practical effects of applicant reactions: Subgroup differences in test-taking motivation, test performance, and selection rates. *International Journal Of Selection And Assessment*, 10(4), 258-270.
- Ployhart, R. E., Schneider, B., & Schmitt, N. (2006). *Staffing organizations: Contemporary practice and theory*, 3rd ed. Mahwah, NJ US: Lawrence Erlbaum Associates Publishers.
- Posthuma, R. A. (2002). Employee selection procedures and the business necessity defense. *Applied H.R.M. Research*, 7(1-2), 53-63.
- Pyburn, K. r., Ployhart, R. E., & Kravitz, D. A. (2008). The diversity-validity dilemma: Overview and legal context. *Personnel Psychology*, 61(1), 143-151.
- Quaintance, M. K. (1980). The impact of the Uniform Selection Guidelines on public merit systems. *Public Personnel Management*, 9(3), 125-133.
- Reilly, R. R. (1996). Alternative selection procedures. In R. S. Barrett (Ed.) , *Fair employment strategies in human resource management* (pp. 208-221). Westport, CT US: Quorum Books/Greenwood Publishing Group.

- Reynolds, D. H., & Knapp, D. J. (2011). SIOP as advocate: Developing a platform for action. *Industrial And Organizational Psychology: Perspectives On Science And Practice*, 4(4), 540-544.
- Risavy, S. D., & Hausdorf, P. A. (2011). Personality testing in personnel selection: Adverse impact and differential hiring rates. *International Journal Of Selection And Assessment*, 19(1), 18-30.
- Ross, R. R., & Altmaier, E. M. (1990). Job analysis of psychology internships in counseling center settings. *Journal Of Counseling Psychology*, 37(4), 459-464.
- Roth, P. L., Bobko, P., & Switzer, F. (2006). Modeling the behavior of the 4/5ths rule for determining adverse impact: Reasons for caution. *Journal Of Applied Psychology*, 91(3), 507-522.
- Roth, P. L., Huffcutt, A. I., & Bobko, P. (2003). Ethnic group differences in measures of job performance: A new meta-analysis. *Journal Of Applied Psychology*, 88(4), 694-706.
- Saal, F. E., & Moore, S. (1993). Perceptions of promotion fairness and promotion candidates' qualifications. *Journal Of Applied Psychology*, 78(1), 105-110.
- Sackett, P. R. (2011). The Uniform Guidelines is not a scientific document: Implications for expert testimony. *Industrial And Organizational Psychology: Perspectives On Science And Practice*, 4(4), 545-546.
- Sackett, P. R. (2012). Cognitive tests, constructs, and content validity: A commentary on Schmidt (2012). *International Journal Of Selection And Assessment*, 20(1), 24-27.
- Satterwhite, R. C., Fleenor, J. W., Braddy, P. W., Feldman, J., & Hoopes, L. (2009). A case for homogeneity of personality at the occupational level. *International Journal Of Selection And Assessment*, 17(2), 154-164.
- Schmidt, F. L. (2012). Cognitive tests used in selection can have content validity as well as criterion validity: A broader research review and implications for practice. *International Journal Of Selection And Assessment*, 20(1), 1-13.
- Schmidt, F. L. (2012). Content validity and cognitive tests: Response to Kehoe (2012), Ployhart (2012), and Sackett (2012). *International Journal Of Selection And Assessment*, 20(1), 28-35.
- Schmidt, F. L., Hunter, J. E., & Pearlman, K. (1981). Correction to Schmidt, Hunter, and Pearlman. *Journal Of Applied Psychology*, 66(6).
- Schmidt, F. L., Hunter, J. E., & Pearlman, K. (1981). Task differences as moderators of aptitude test validity in selection: A red herring. *Journal Of Applied Psychology*, 66(2), 166-185.
- Sharf, J. C. (2011). Equal employment versus equal opportunity: A naked political agenda covered by a scientific fig leaf. *Industrial And Organizational Psychology: Perspectives On Science And Practice*, 4(4), 537-539.

Sireci, S. G., & Geisinger, K. F. (1998). Equity issues in employment testing. In J. H. Sandoval, C. L. Frisby, K. F. Geisinger, J. Scheuneman, J. Grenier (Eds.), *Test interpretation and diversity: Achieving equity in assessment* (pp. 105-140). Washington, DC US: American Psychological Association.

Suen, H. K., & French, J. L. (2003). A history of the development of psychological and educational testing. In C. R. Reynolds, R. W. Kamphaus (Eds.), *Handbook of psychological and educational assessment of children: Intelligence, aptitude, and achievement* (2nd ed.) (pp. 3-23). New York, NY US: Guilford Press.

Thayer, P. W. (1983). Industrial/organizational psychology: Science and application. In C. Scheirer, A. M. Rogers (Eds.) *The G. Stanley Hall lecture series*, Vol. 3 (pp. 9-32). Washington, DC US: American Psychological Association.

Thornton, G., & Mueller-Hanson, R. A. (2004). *Developing organizational simulations: A guide for practitioners and students*. Mahwah, NJ US: Lawrence Erlbaum Associates Publishers.

Thornton, G., & Rupp, D. E. (2004). Simulations and assessment centers. In J. C. Thomas (Ed.) *Comprehensive handbook of psychological assessment*, Vol. 4: Industrial and organizational assessment (pp. 319-344). Hoboken, NJ US: John Wiley & Sons Inc.

Thornton, G., & Rupp, D. E. (2006). *Assessment centers in human resource management: Strategies for prediction, diagnosis, and development*. Mahwah, NJ US: Lawrence Erlbaum Associates Publishers.

Tippins, N. T. (2013). Assessment of leadership. In K. F. Geisinger, B. A. Bracken, J. F. Carlson, J. C. Hansen, N. R. Kuncel, S. P. Reise, M. C. Rodriguez (Eds.) *APA handbook of testing and assessment in psychology*, Vol. 1: Test theory and testing and assessment in industrial and organizational psychology (pp. 457-478). Washington, DC US: American Psychological Association.

Tippins, N. T., & Macey, W. H. (2007). Consortium Studies. In S. McPhail (Ed.), *Alternative validation strategies: Developing new and leveraging existing validity evidence* (pp. 233-251). Hoboken, NJ US: John Wiley & Sons Inc.

Tippins, N. T., Papinchock, J. M., & Solberg, E. C. (2010). Decisions in developing and selecting assessment tools. In J. L. Farr, N. T. Tippins (Eds.), *Handbook of employee selection* (pp. 363-376). New York, NY US: Routledge/Taylor & Francis Group.

Tonowski, R. F. (2011). The Uniform Guidelines and personnel selection: Identify and fix the right problem. *Industrial And Organizational Psychology: Perspectives On Science And Practice*, 4(4), 521-525.

Trattner, M. H. (1982). Synthetic validity and its application to the Uniform Guidelines validation requirements. *Personnel Psychology*, 35(2), 383-397.

Van Scotter, J. R., Moustafa, K., Burnett, J. R., & Michael, P. G. (2007). Influence of prior acquaintance with the ratee on rater accuracy and halo. *Journal Of Management Development*, 26(8), 790-803.

Veres, J. G., Lahey, M. A., & Buckley, R. (1987). A practical rationale for using multi-method job analyses. *Public Personnel Management*, 16(2), 153-157.

Vinchur, A. J. (2007). A History of Psychology Applied to Employee Selection. In L. L. Koppes (Ed.) , *Historical perspectives in industrial and organizational psychology* (pp. 193-218). Mahwah, NJ US: Lawrence Erlbaum Associates Publishers.

Wang, S., Wang, N., & Hoadley, D. (2007). Construct equivalence of a national certification examination that uses dual languages and audio assistance. *International Journal Of Testing*, 7(3), 255-268.

Weitzman, R. A. (1991). 'Unfairness' in the 1978 Uniform Guidelines: The definition makes a difference. *Psychological Reports*, 69(2), 471-482.

Wingate, P. H., Thornton, G., McIntyre, K. S., & Frame, J. H. (2003). Organizational downsizing and age discrimination litigation: The influence of personnel practices and statistical evidence on litigation outcomes. *Law And Human Behavior*, 27(1), 87-108.

References Citing Barrett and Kernan (1987)

- Austin, J.T., & Villanova, P. (1992). The criterion Problem 1917-1992. *Journal of Applied Psychology*, 77, 836-874.
- Barrett, G. V. (2000). Personnel selection: Selection and the law. In A. E. Kazdin (Ed.), *Encyclopedia of psychology*, Vol. 6 (pp. 156-160). Washington, DC New York, NY USUS: American Psychological Association.
- Bernardin, H., Hennessey, H. r., & Peyrefitte, J. (1995). Age, racial, and gender bias as a function criterion specificity: A test of expert testimony. *Human Resource Management Review*, 5(1), 63-77.
- Binning, J. F., & Barrett, G. V. (1989). Validity of personnel decisions: A conceptual analysis of the inferential and evidential bases. *Journal Of Applied Psychology*, 74(3), 478-494.
- Bretz, R.D., Milkovich, G.T., & Read, W. (1992). The current state of performance-appraisal research and practice - Concerns, directions, and implications. *Journal of Management*, 18, 321-352.
- Cascio, W.F., Alexander, R.A., & Barrett, G.V. (1988). Setting cutoff scores - Legal, psychometric, and professional issues and guidelines. *Personnel Psychology*, 41, 1-24.
- Cesare, S.J., Blankenship, M.H., Giannetto, P.W., & Mandel, M.Z. (1993). A predictive validation-study of the methods used to select eligibility technicians. *Public Personnel Management*, 22, 107-122.
- Commisso, M., & Finkelstein, L. (2012). Physical attractiveness bias in employee termination. *Journal Of Applied Social Psychology*, 42(12), 2968-2987.
- Costello, R.M., Schneider, S.L., & Schoenfeld, L.S. (1996). Validation of a preemployment MMPI index correlated with disciplinary suspension days of police officers. *Psychology Crime & Law*, 2, 299-306.
- Eberlin, R.J., & Tatum, B. (2005). Conceptual paper: Organizational justice and decision making: When good intentions are not enough. *Management Decision*, 43(7-8), 1040-1048.
- Eberlin, R. J., & Tatum, B. (2008). Making just decisions: Organizational justice, decision making, and leadership. *Management Decision*, 46(2), 310-329.
- Elkins, T. J., & Philips, J. S. (1999). Evaluating sex discrimination claims: The mediating role of attributions. *Journal Of Applied Psychology*, 84(2), 186-199.
- Elkins, T. J., Phillips, J. S., Konopaske, R., & Townsend, J. (2001). Evaluating gender discrimination claims: Is there a gender similarity bias? *Sex Roles*, 44(1-2), 1-15.
- Fletcher, C., & Perry, E. L. (2002). Performance appraisal and feedback: A consideration of national culture and a review of contemporary research and future trends. In N. Anderson, D. S.

Ones, H. Sinangil, C. Viswesvaran (Eds.) , Handbook of industrial, work and organizational psychology, Volume 1: Personnel psychology (pp. 127-144). Thousand Oaks, CA: Sage Publications Ltd.

Harris, N.M., Smith, D.E., & Champagne, D. (1995). A field-study of performance-appraisal purpose - Research-based versus administrative-based ratings. *Personnel Psychology*, 48, 151-160.

Hedge, J. W., Borman, W. C., & Lammlein, S. E. (2006). The aging workforce: Realities, myths, and implications for organizations. Washington, DC US: American Psychological Association.

Komaki, J. L. (2007). Daring to dream: Promoting social and economic justice at work. *Applied Psychology: An International Review*, 56(4), 624-662.

Martin, D.C., & Bartol, K.M. (1991). The legal ramifications of performance-appraisal - An update. *Employee Relations Law Journal*, 17, 257-286.

Maurer, T.J., & Alexander, R.A., (1992). Methods of improving employment test critical scores derived by judging test content - A review and critique. *Personnel Psychology*, 45, 727-762.

Miller, C.S., Kaspin, J.A., & Schuster, M.H. (1990). The impact of performance-appraisal methods on age-discrimination in employment act cases. *Personnel Psychology*, 43, 555-578.

Murphy, K. R., & Deckert, P. J. (2013). Performance appraisal. In K. F. Geisinger, B. A. Bracken, J. F. Carlson, J. C. Hansen, N. R. Kuncel, S. P. Reise, M. C. Rodriguez (Eds.) , *APA handbook of testing and assessment in psychology*, Vol. 1: Test theory and testing and assessment in industrial and organizational psychology (pp. 611-627). Washington, DC US: American Psychological Association.

Newman, D. A., Kinney, T., & Farr, J. L. (2004). Job performance ratings. In J. C. Thomas (Ed.) , *Comprehensive handbook of psychological assessment*, Vol. 4: Industrial and organizational assessment (pp. 373-389). Hoboken, NJ US: John Wiley & Sons Inc.

Organ, D. W., Podsakoff, P. M., & Podsakoff, N. P. (2011). Expanding the criterion domain to include organizational citizenship behavior: Implications for employee selection. In S. Zedeck (Ed.) , *APA handbook of industrial and organizational psychology*, Vol 2: Selecting and developing members for the organization (pp. 281-323). Washington, DC US: American Psychological Association.

Phillips, S.D., Cairo, P.C., Blustien, D.L. & Myers R.A. (1998). Career-development and vocational behavior, 1987 a review. *Journal of Vocational Behavior*, 33, 119-184.

Redman, T., & Snape, E. (1992). Upward and onward - Can staff appraise their managers? *Personnel Review*, 21, 32-46.

Robinson, R.K., Fink, R.L., & Allen, B.M. (1996). The influence of organizational constituent groups on rater attitudes toward performance appraisal compliance. *Public Personnel Management*, 25, 141-150.

Roehling, M.V. (1993). Extracting policy from judicial opinions - The dangers of policy capturing in a field setting. *Personnel Psychology*, 46, 477-502.

Scholl, R.W., & Cooper, E. (1991). The use of job evaluation to eliminate gender based pay differentials. *Public Personnel Management*, 20, 1-18.

Schmitt, N., & Robertson, I. (1990). Personnel Selection. *Annual Review of Psychology*, 41, 289-319.

Snyder, C.J., & Barrett, G.V. (1988). The Age-discrimination in employment act - A review of court decisions. *Experimental Aging Research*, 14, 3-47.

Swanson, D. R. (2002). Diversity programs: Attitude and realities in the contemporary corporate environment. *Corporate Communications*, 7(4), 257-268.

Tatum, B., & Eberlin, R. J. (2006). Organizational justice and conflict management styles: Teaching notes, role playing instructions, and scenarios. *International Journal Of Conflict Management*, 17(1), 66-81.

Werner, JM, & Bolino, M.C. (1997). Explaining US courts of appeals decisions involving performance appraisal: Accuracy, fairness, and validation. *Personnel Psychology*, 50, 1-24.

Wingate, P. H., Thornton, G., McIntyre, K. S., & Frame, J. H. (2003). Organizational downsizing and age discrimination litigation: The influence of personnel practices and statistical evidence on litigation outcomes. *Law And Human Behavior*, 27(1), 87-108.